



xx

COPY RIGHT

2024 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 10th july 2024. Link

<https://www.ijiemr.org/downloads/Volume-13/Issue-7>
10.48047/IJIEMR/V13/ISSUE 07/04

TITLE: OPTIMISING DATA PERFORMANCE: TECHNIQUES FOR FASTER AND EFFICIENT DATA PROCESSING

Volume 13, ISSUE 07, Pages: 25-27

Paper Authors : SAI CHARAN DATTA



USE THIS BARCODE ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code

OPTIMISING DATA PERFORMANCE: TECHNIQUES FOR FASTER AND EFFICIENT DATA PROCESSING

AUTHOR - SAI CHARAN DATTA

EMAIL – CHARAN.DATTA@GMAIL.COM

COMPANY – SYSTEM SOFT TECHNOLOGIES

Abstract— Context: Large set of data creates challenges for businesses to process it for organisational performance improvement activities. Increasing cost and ineffective processing activities has created challenges for optimised data performance management.

Objective: Main objective of this research is to identify effective techniques for faster and efficient data processing to optimise data performance.

Method: A secondary qualitative data collection process along with thematic analysis techniques assist to develop credible research findings on data process techniques and optimisation of data performance.

Results: The result section of this study has identified the appropriate techniques which are beneficial in fostering time and efficacy in data processing. It has been found that the techniques like parallel processing, optimisation algorithms, cache utilisation are beneficial in reducing time in data processing.

Conclusion: Data performance help in fostering the quality of quality and positively influence business operations.

Keywords: Data performance, data processing, time-consumption, efficiency

I. INTRODUCTION

Data optimisation helps businesses to extract, store and evaluate data to the maximum efficiency. Optimisation of data performance assists organisations to manage workload of operation, computation activities and transactional processes that depend on accurate retrieval of data, its processes and storage management [1]. Modern businesses depending on open source data format collection have assisted to collect information according to format needs of each application in an organisation. This open source data utilisation activities creates columnar storage blockage in data utilisation that were collected in open source format. In this context, this research has focused on developing a critical viewpoint on optimisation of data performance. This research would be also assessing different techniques for faster as well as efficient data processes activities for improving organisational capabilities in resource effectiveness management and enhancing decision making scope.

A) Aim and Research Objectives

Aim

The aim of this research paper is to develop critical knowledge about optimisation of data performance while evaluating effective techniques for faster and efficient data processing activities.

Objectives

- To investigate importance of optimising data performance
- To identify effective techniques for faster and efficient data processing to optimise data performance
- To evaluate challenges of faster and efficient data processing for optimised data performance management
- To recommended effective strategies for faster and efficient data processing to optimise data performance

B) Background

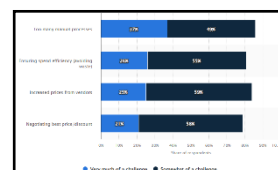


Figure 1: “Challenges in optimising IT spend in global companies in 2022”

(Source: [2])

High volume of data increases the cost associated with relying on data for making appropriate decisions for improving organisational performance management capabilities. Approximately 84% of respondents have identified increased price from vendors creates challenges for data optimisation system development due to huge data repository management [2]. Besides, management of manual processes for data processing, speed efficiency management and negotiation of best prices of discounts are other major challenges related to optimising data performance. Dealing with massive amounts of data creates challenges for businesses in organised management of internal activities while managing resource capabilities and decision making approaches. Ineffective data optimisation activities creates challenges for businesses to prioritise change management due to unreliable information collection creates difficulties to predict effectively [3]. Besides, data optimisation issues would create challenges for businesses in resource effectiveness management while improving organisational capabilities for effective decision making activities to improve performance. Hence, lack of optimisation of data performance systems has been affecting businesses to manage systematic approaches of operational planning, resource effectiveness and decision making activities. Considering it, this research has focused on critically evaluating different data processing techniques for faster and efficient data assessment and to optimise data performance.

II. LITERATURE REVIEW

Importance of optimising data performance

Data optimisation improving overall data quality ensures meaningful data driven decision making activities. Effective data quality management activities associated with data processing, profiling and cleansing processes for using the data for operation management [4]. Data optimisation activities recognising internal processing requirements of data while explicitly assessing differences between average values would be beneficial for improving overall data quality for businesses. In association with improving data quality, optimisation of data performance helps in enhancing data monitoring activities. Efficient data processing module managing real time data measuring activities helps to manage data transmission, calculation and statistical processing [5]. Effective data monitoring systems changing accordingly to meet requirements of diverse sites, communication protocols and environmental protection bureaus ensure efficiency in functional structure of optimising data processing activities. Hence, effective optimisation of data performance improving data quality and monitoring activities benefits to enhance operational effectiveness of businesses.

Techniques for faster and efficient data processing to optimise data performance

Data mart can be identified as an effective technique for businesses to process a small amount of data for extracting useful information. Data mart techniques easily customise the collected data to sort, select, structure and summarise the

same to meet requirements of a department [6]. This data mart technique gathers data from external sources and stores them at a data warehouse for managing systematic approaches of data processes activities. Further, effective use of data mart techniques improves organisational capabilities to effectively manage technological and economic aspects help in data optimisation management. Use of progressive learning techniques such as DenseNet and EfficientNet helps on improving parameter efficiency to achieve better accuracy in data processing activities [7]. These techniques also aim to enhance overall training and inference speed management in data processes activities along with parameter efficiency management practices. This helps to decrease the cost of efficiency and improve training efficiency for overall accuracy, efficiency and speed of data processing activities. Thus, data mart and progressive learning techniques benefit in efficient and faster data processes activities to optimise data performance.

Challenges of faster and efficient data processing to manage optimised data performance

Lack of effective quality management of data has created challenges for faster and efficient data processing activities to manage optimised data performance. Inconsistencies in data along with data duplication and incomplete information create issues regarding effective data quality management [8]. Further, ineffective data quality management creates challenges for comparative analysis, measurement and overall data processing activities for development of optimised data performance. On the other hand, exponential data volume growth has created serious challenges for computing, processing and analysing large scales of information [9]. Massive data volume would also create challenges related to data collecting, storage layering and in-depth analysis activities. Therefore, large amounts of data volume and data quality issues create challenges for faster and efficient data processes management for optimising data performance.

Strategies of faster and efficient data processing to optimise data performance
Integration of effective strategies such as parallel processing strategies benefits in faster and efficient data processes management to optimise overall data performance. Parallel processing activities run N (number-of-processors) times faster than could be done on a single processor, based on "Processor Utilisation Efficiency (PUE)" as ratio of single processor time to parallel processor time divided by N [10]. These processing applications split independent elements to run in parallel on a huge number of processors to improve efficiency in data processing and optimise data performance. Optimisation of algorithms can be identified as another data processing approach for optimising data performance. Optimization Algorithm processes uses meta-heuristics methods for improvement of robustness, performance reliability, simplicity, ease of implementation [11]. Optimised algorithm processes also developing evolutionary-based, swarm-based, physics-based, human-behaviour based and hybrid algorithm development assist to manage subjective nature in data processing approaches to optimise data performance.

Theoretical Perspectives

Application of SVM Model helps to solve classification and regression tasks particularly solving binary classification problems for classification of data. SVM models use optimal parameters for direct evaluation of fit and performance of data [12]. This model helping assessment of higher dimension of samples has created opportunities for efficient and optimised data performance management. Accordingly, effective use of SVM model benefits in improving speed and efficiency in data processing activities for optimisation of data performance.

Gap in Literature

A major in literature was identified in terms of developing clear knowledge about data processing techniques failing to develop optimise data performance. Lack of information about effectiveness of data processes techniques has limited knowledge development of its abilities for optimisation of data performance. Following it, this research would focus on developing critical knowledge about faster and efficient data processing techniques to optimise data performance.

III. METHODOLOGY

Systematic development of research methodology guides to develop organised techniques for data collection and evaluation activities. This research has followed "Saunders research onion" for identification of appropriate philosophy, approach, strategy, data collection and data analysis techniques. "Saunders research onion" illustrates systematic layers for data collection and analysis techniques [13]. Considering it, this research has used "interpretivism

research philosophy" to manage subjective nature in data collection activities. Besides, "inductive approach" has assisted this research to develop critical knowledge about data processing activities optimising data performance. Further, effective use of "archival research strategy" has assisted this research to gather a wide range of information from already published journal articles.

This research has followed "secondary qualitative data collection" processes for gathering information about techniques for faster and efficient data processing for optimising data performance. "Secondary data collection" helps in time and cost effective information collection processes on research variables [14]. Secondary data collection processes collect information from peer reviewed sources such as "IEEE access", "PHI Learning Pvt. Ltd" and "Journal of Internet Technology". These peer reviewed journals have assisted to gather massive amounts of information about data processing techniques and optimise data performance management activities.

"Thematic analysis" technique was used in this research for analysing and representing collected data in a systematic manner. A total of 4 themes were developed following research objectives to represent collected data in a systematic manner. Thematic analysis helps in comprehensive research findings development by critical assessment of complex data [15]. Effective use of thematic analysis has assisted this research to develop sensible data findings on techniques to use for faster and efficient data processing activities for optimising data performance. Thus, secondary qualitative data analysed through thematic analysis techniques has assisted this research for subjective and credible data findings development activities on optimisation of data performance.

IV. RESULTS AND DISCUSSION

A) Result

Theme 1: Optimising data performance has helped in improving data quality and data processing

Data optimisations refers to the process of getting accurate and reliable data along with cleaning the error data. The activity of data performance optimization has become important for improving the quality of data and making it usable for data analysis, decision-making along with other business operations [16]. Accordingly, it can be possible to transform the raw data into relevant, uniform along with updated data with the application of data optimisation. This further contributes to maximising the use and application of data. On the other hand, data performance optimization is beneficial in leading toward faster data processing, improvement in user experience along with scalability of systems [17]. Data performance optimization works on laying the foundation for efficient data monitoring along with ensuring accuracy of the insights gained from refined data analysis. Thus, it becomes beneficial to ensure effective data processing with data optimisation along with leading effective decisions making.

Theme 2: Parallel processing and optimisation algorithms have fostered efficient data processing

Data performance optimization and data processing is dependent on a range of techniques including optimization algorithms, fixing data errors along with the use of the advanced technology. The integration of parallel processing has become one of the significant techniques in ensuring faster data processing as it works on distributing data processing tasks across multiple nodes [18]. Accordingly, the application of parallel processing helps in executing the process of data performance optimization. The application of parallel processing thus contributes to reducing the time of processing along with fostering resource utilisation. On the contrary, optimisation algorithms are considered as mathematical techniques which help in finding the exact solution for any given problem [19]. Optimisation algorithms help in leading toward the reduction of resource consumption through minimising the use of system resources including disk space, CPU and memory. As a result, the utilisation of techniques like optimisation algorithms become beneficial for supporting faster and efficient data processing.

Theme 3: The issues associated with data quality, data integration and inconsistency of data have created challenge for data processing and data performance optimisation

Data processing encompasses a number of challenges which are the issues in data collection, data inconsistency along with duplication of data. The collection of the appropriate and enormous amount of data has become one of the major challenges in data processing [20]. Data processing requires correct data for producing validated results and lack of data creates difficulties in leading toward faster data processing. Conversely, duplication of data is another

issue which results in creating errors at the time of data processing [21]. In this context, the collection of the data from multiple sources results in multiple entries of the same information. This further contributes to creating challenges at the stage of data encoding through producing incorrect results. Therefore, the presence of duplicate data within the collected data creates a barrier to efficient data processing.

Theme 4: Optimising data pipelines and cache utilisation can help in maximising the efficiency of data processing and performance

The integration of the strategies like optimising data pipelines and cache utilisation come with the opportunity to the business in leading toward efficient data processing. Optimising data pipelines includes the activities of minimising the errors and issues in data for ensuring smooth data flow [22]. In this context, the businesses can work on recognising and resolving unnecessary steps along with prioritising critical transformations for maximising optimisation in data pipelines. However, cache utilisation is another significant strategy which helps in delivering temporary storage solutions for faster data processing [23]. The advanced mechanism of cache utilisation delivers the opportunities of reducing the need for reloading and computing data. Henceforth, cache utilisation contributes to making the data processing much easier and faster along with optimising data performance.

B) Discussion

The result of thematic analysis has identified the importance of optimising data performance along with recognising the techniques for faster data processing. Data optimisation has found to be beneficial in improving the quality of data along with helping the streamline flow of the operations of the businesses. Similarly, the findings from the literature review have also mentioned about the significance of optimising data performance with the application of the appropriate tools for automating the process along with reducing the time associated with data processing [6]. The techniques that are identified as beneficial for efficient and faster data processing includes optimisation algorithms and parallel processing. However, the literature review has mentioned techniques like data mart which become beneficial in leading appropriate data processing. Thus, the findings of thematic analysis and literature review have discussed about the tools and techniques for contributing to the data processing.

V. CONCLUSION

This study concludes the importance of optimising data performance and the used techniques for faster data processing. It has been found that data performance optimisation results in ensuring effective use of the resources, faster data processing along with minimising error in data. Furthermore, this study has identified the techniques associated with faster data processing which are parallel processing, data compression, cache utilisation and others. The use of the secondary data collection and thematic analysis are found as beneficial for gathering large volumes of information associated with data performance and data processing. This study has summarised the importance of data performance optimisation in improving workflow, work efficiency and effective communication within the workplaces. The strategies for improving data processing are recognised in this study which includes the activities of optimising data pipelines.

The strategies that can be recommended for further improving data processing includes Extract, transform, and load (ETL) optimisation. Application of ETL technique comes with the opportunity of minimising processing time through streamlining the process of extraction along with implementing efficient transformation algorithms [24]. The activities of the businesses in making each stage efficient help in automatically influencing workflow and data performance optimisation. Therefore, it can be recommended to the business to focus on the use of the ETL technique for optimising the activities of data processing and loading process.

VI. ACKNOWLEDGEMENT

I would like to thank my family members and parents for supporting me in completing the paper. I am thankful to my professors for assisting me.

REFERENCES

[1] Costa, R.L.D.C., Moreira, J., Pintor, P., dos Santos, V. and Lifschitz, S., (2021). A survey on data-driven performance tuning for big data analytics platforms. *Big Data Research*, 25, p.100206.

[2] Statista.com (2024). IT spend optimization challenges worldwide 2022. Available at: <https://www.statista.com/statistics/1106051/challenges-in-managing-it-spend/>. [Accessed on: 15 June 2024]

[3] Nudurupati, S.S., Tebboune, S., Garengo, P., Daley, R. and Hardman, J., (2024). Performance measurement in data intensive organisations: resources and capabilities for decision-making process. *Production Planning & Control*, 35(4), pp.373-393.

[4] Ehrlinger, L. and WöB, W., (2022). A survey of data quality measurement and monitoring tools. *Frontiers in big data*, 5, p.850611.

[5] Wang, B., Yao, X., Jiang, Y., Sun, C. and Shabaz, M., (2021). Design of a real-time monitoring system for smoke and dust in thermal power plants based on improved genetic algorithm. *Journal of Healthcare Engineering*, 2021(1), p.7212567.

[6] Prabhu, C.S.R., (2008). *Data warehousing: concepts, techniques, products and applications*. PHI Learning Pvt. Ltd.

[7] Tan, M. and Le, Q., (2021, July). Efficientnetv2: Smaller models and faster training. In *International conference on machine learning* (pp. 10096-10106). PMLR.

[8] Fan, W. and Geerts, F., (2022). *Foundations of data quality management*. Springer Nature.

[9] Wang, J., Yang, Y., Wang, T., Sherratt, R.S. and Zhang, J., (2020). Big data service architecture: a survey. *Journal of Internet Technology*, 21(2), pp.393-405.

[10] Cave, W.C., Wassmer, R.E., Ledgard, H.F., Salisbury, A.B., Irvine, K.T. and Mulshine, M.A., (2020). A new approach to parallel processing. *IEEE Access*, 8, pp.30287-30305.

[11] Ayyarao, T.S., Ramakrishna, N.S.S., Elavarasan, R.M., Polumahanthi, N., Rambabu, M., Saini, G., Khan, B. and Alatas, B., 2022. War strategy optimization algorithm: a new effective metaheuristic algorithm for global optimization. *IEEE Access*, 10, pp.25073-25105.

[12] Zhao, S. and Zhao, Z., (2021). A Comparative Study of Landslide Susceptibility Mapping Using SVM and PSO-SVM Models Based on Grid and Slope Units. *Mathematical problems in Engineering*, 2021(1), p.8854606.

[13] Saunders M., Thornhill, A. and Lewis P. (2019) *Methods for Business Students* (8th edition). Person Education: London.

[14] Taherdoost, H., (2021). Data collection methods and tools for research; a step-by-step guide to choose data collection technique for academic and business research projects. *International Journal of Academic Research in Management (IJARM)*, 10(1), pp.10-38.

[15] Wiltshire, G. and Ronkainen, N., (2021). A realist approach to thematic analysis: making sense of qualitative data through experiential, inferential and dispositional themes. *Journal of Critical Realism*, 20(2), pp.159-180.

[16] Hassan, C.A.U., Hammad, M., Uddin, M., Iqbal, J., Sahi, J., Hussain, S. and Ullah, S.S., (2022). Optimizing the performance of data warehouse by query cache mechanism. *IEEE Access*, 10, pp.13472-13480.

[17] Naeem, M.A., Mirza, F., Khan, H.U., Sundaram, D., Jamil, N. and Weber, G., (2020). Big data velocity management—from stream to warehouse via high performance memory optimized index join. *IEEE Access*, 8, pp.195370-195384.

[18] Cave, W.C., Wassmer, R.E., Ledgard, H.F., Salisbury, A.B., Irvine, K.T. and Mulshine, M.A., (2020). A new approach to parallel processing. *IEEE Access*, 8, pp.30287-30305.

[19] Oyelade, O.N., Ezugwu, A.E.S., Mohamed, T.I. and Abualigah, L., (2022). Ebola optimization search algorithm: A new nature-inspired metaheuristic optimization algorithm. *IEEE Access*, 10, pp.16150-16177.

[20] Bhat, S.A. and Huang, N.F., (2021). Big data and ai revolution in precision agriculture: Survey and challenges. *Ieee Access*, 9, pp.110209-110222.

[21] Pandey, A.K., Khan, A.I., Abushark, Y.B., Alam, M.M., Agrawal, A., Kumar, R. and Khan, R.A., (2020). Key issues in healthcare data integrity: Analysis and recommendations. *IEEE Access*, 8, pp.40612-40628.

[22] Nagorny, K., Scholze, S., Colombo, A.W. and Oliveira, J.B., (2020). A DIN Spec 91345 RAMI 4.0 compliant data pipelining model: An approach to support data understanding and data acquisition in smart manufacturing environments. *IEEE Access*, 8, pp.223114-223129.

[23] Denzler, A., Oliveira, G.F., Hajinazar, N., Bera, R., Singh, G., Gómez-Luna, J. and Mutlu, O., (2023). Casper: Accelerating Stencil Computations Using Near-Cache Processing. *IEEE Access*, 11, pp.22136-22154.

[24] Mehmood, E. and Anees, T., (2020). Challenges and solutions for processing real-time big data stream: a systematic literature review. *IEEE Access*, 8, pp.119123-119143.